DOCUMENT RESUME

ED 302 770 CG 021 359

AUTHOR Steiner, Dirk D.; Rain, Jeffrey S.
TITLE Rating Variable Performance.
PUB DATE 14 Aug 88
NOTE 24p.; Paper presented at the Annual Meeting of the American Psychological Association (96th, Atlanta, GA, August 12-16, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Behavior Patterns; College Students; Higher Education; Interrater Reliability; *Measurement Techniques; *Performance Factors; *Rating Scales

ABSTRACT

Many empirical studies have examined factors that influence ratings of performance. This study examined the rating variable performance of a single individual. Serial position of a single poor or good performance in a series of otherwise good or poor performances was manipulated to examine its effects on both ratings and recommended actions toward the ratee. Undergraduate students (N=564) viewed four videotaped lectures either in one session or over 4 days. Behavioral Observation Scale (BOS) ratings of performance across the four lectures were unaffected by a single poor performance in a series of good performances. Overall ratings on a one-item scale showed greater effects. In the single session conditions, a recency effect resulted such that the overall rating was given in the direction of the most recent performance. In the 4-day sessions, a single good performance did not elevate ratings of poor base performance; but a single good performance may have made establishment of a schema difficult and lowered ratings of good base performance. Similar results were also obtained for recommendations to punish the instructor. (ABL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED302770 -

Rating Variable Performance

Dirk D. Steiner and Jeffrey S. Rain

Louisiana State University

CG 021359

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dirk D. Steiner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 96th Annual Convention of the American Psychological Association. Atlanta, August 14, 1988.

BEST COPY AVAILABLE

Abstract

Serial position of a single poor or good performance in a series of otherwise good or poor performances was manipulated to examine its effects on both ratings and recommended actions toward the ratee. 564 undergraduate Ss viewed 4 videotaped lectures in 1 session or over 4 days. Behavioral Observation Scale (BOS) ratings of performance across the 4 lectures were unaffected by a single poor performance in a series of good performances. BOS ratings were higher when a single good performance occurred in later positions in a series of poor performances. Overall ratings on a 1-item scale showed greater effects. In the single session conditions, a recency effect resulted such that the overall rating was given in the direction of the most recent performance. In the 4-day sessions, a single good performance did not elevate ratings of poor base performance; but a single poor performance may have made establishment of a schema difficult and lowered ratings of good base performance. Similar results were also obtained for recommendations to punish the instructor.

Rating Variable Performance

At the beginning of this decade, studies of performance appraisal shifted focus from the study of rating forms to the examination of cognitive processes (Cooper, 1981; DeNisi, Cafferty, & Meglino, 1984; Feldman, 1981; Landy & Farr, 1980). As a result, a rater is depicted as observing behavior, storing the observations in memory after processing them, recalling the stored information at a later time, and then translating the recalled information into some judgement of the performance.

As a result of this shift in focus, many empirical studies have examined factors that influence ratings of performance. Much of this research assumes that performance is stable. In fact, if raters evaluate performance differently over time, the variation is usually attributed to unreliability of the rater (Landy & Farr, 1983). There is a body of research, however, which indicates that individual performance varies over time (Rambo, Chomiak, & Price, 1983; Ronan & Prien, 1971; Rothe, 1978). Kane and Lawler (1979) also note that periods of coasting or bursts of achievement are commonly observed in people. While the recent cognitive approaches do include implications for rating variable performance (cf. Cooper, 1981; DeNisi et al., 1984), to date, little research has been done to address how raters process variable or inconsistent information about a single individual.

Murphy and his colleagues have conducted research that indirectly addresses how raters process variable information in ratee performance. Their research studied perceptions of specific incidents of performance. In 1985, Murphy, Balzer, Lockhart, and Eisenman reported a contrast-effect bias

such that a ratee's recent average performance was rated lower if the same ratee's previous performance was good and higher if previous performance was poor. The effect was not present when memory demands increased, and Murphy et al. concluded that the contrast effect was due to greater attention to and richer encoding of inconsistent performance. In a second study, Murphy, Gannett, Herr, and Chen (1986) found an assimilation effect. An assimilation effect operates in the opposite direction of a contrast effect by making perceptions of inconsistent information appear similar to the other information. In this study it was demonstrated by recall of the ratee's previous performance being biased in the direction of the ratee's more recent performance, but only with increased memory demands. Murphy et al. attributed this assimilation effect to the development of a schema which biases memory for ratee behavior. Based on both studies, Murphy et al. (1986) concluded that contrast effects result in conditions where attentional processes are maximized and memory demands are minimized, while assimilation effects result when attention is minimized and memory demands are great. These studies, however, did not require raters to integrate the several observations into a single overall impression. Rather their focus was on the rating of a particular observation in the context of other observations. An integrated, overall rating of an individual's performance is required in typical organizational appraisal situations, even when employee performance is variable.

A study by Scott and Hamner (1975) did obtain overall ratings. They found no effect for performance variability or increasing or decreasing performance patterns on ratings of overall performance. Subjects in this

study, however, rated performance on a marble-bagging task for which the objective criterion of quantity of performance was readily observed. DeNisi and Stevens (1981) found that variable performance was rated lower than stable performance, except at the high performance level where variability had no effect. A more recent study by Steiner and Rain (in press) asked subjects to provide overall evaluations of an instructor based on viewing four short videotapes of his teaching. Three of the excerpts represented average performance, while a fourth, occurring in varying serial positions, represented either poor or good performance. The authors found evidence for a recency effect, in that the overall rating was biased in the direction of the inconsistent performance if it was in the last serial position. They concluded that the attention decrement hypothesis best accounted for the recency effect (Anderson, 1971; Luchins, 1957; Schneider, Hastorf, & Ellsworth, 1979). According to this hypothesis, a primacy effect results due to waning attention as people continue to observe while a recency effect occurs when attention is maintained throughout the observation period.

The current study continues the examination of rating variable performance of a single individual. It extends the research of Steiner and Rain on two major points. First, the inconsistent performance for this study differed more from base performance in order to study more clear cases of variability in performance. In Steiner and Rain (in press), the good or poor performance may not have been noticeably different from average performance. And second, subjects in Steiner and Rain (in press) rated each performance excerpt immediately following observation and rated overall performance at the end of the experimental session. This may have helped

maintain attention and contributed to the recency effect. The current study, therefore, only solicited the final overall ratings. The current study was also conducted using immediate and delayed rating conditions as this factor has been found to influence performance ratings significantly (Murphy et al., 1985; Steiner & Rain, in press).

In addition, subjects' recommendations for actions to take toward the instructor were examined.

Method

Procedure Overview. Subjects viewed four videotaped excerpts of lectures by an instructor. Each excerpt was approximately seven minutes in length. The design of the study was a 2 X 2 X 5 factorial. Subjects were run in groups of 5 to 10 which were randomly assigned to one of the experimental cells. There were two levels of time delay, either immediate, in which case subjects viewed all four lectures and rated performance in one one-hour session, or delayed, where subjects viewed one lecture each day for four consecutive days and returned on the fifth day to make their ratings. There were two levels of base performance, good or poor. And there were five serial positions for the presentation of the inconsistent performance. The poor or good inconsistent performance was either omitted (control condition) or occurred first, second, third, or fourth in the series. The same poor lecture was used as inconsistent performance for the good base performance conditions, and the same good one for the poor base performance conditions.

Subjects. 564 undergraduate students, primarily sophomores, at a large southern state university participated in the experiment for extra credit in

their psychology courses. Approximately 28 subjects were assigned to each condition.

Videotapes. A professional actor was hired to portray an instructor giving poor, average, and good quality lectures of each of four lectures. The lectures covered various management topics that students would likely encounter in a first management or industrial psychology course. The actor was given a copy of the rating form (described below) and told to behave in ways consistent with the poor, average, or good ratings on the form. His performance was consistently poor, average, or good across all dimensions. For example, in the poor performance lectures, he acted nervous, spoke in a monotone, and did not vary his facial expression, all behaviors represented by items on the rating form.

Instruments

Rating Form. Ratings were made on a 10-item behavioral observation scale (BOS) for instructor performance adapted from Murphy, Martin, and Garcia (1982). Subjects responded to each item using a 7-point Likert-type scale. Internal consistency reliability was .80 for the scale in past research (Steiner & Rain, in press). Subjects also indicated their overall impression of the instructor using a one-item 7-point scale.

Appropriate Actions. Finally, subjects completed a 24-item instrument regarding how appropriate various actions to take toward the instructor were. The instrument was developed for this study to cover a variety of alternatives for dealing with an instructor. Subjects rated how appropriate each item was on a 9-point scale ranging from not appropriate ("1") to very appropriate ("9"). "Decrease his pay" and "Recommend him for a promotion"

are two items from the instrument. A principal components analysis of the scale yielded three factors by examining eigenvalues and scree plots. An oblique rotation was chosen, and items were retained when their loadings were above .50 on their respective factors. The factors were labelled Counsel, Punish, and Reward.

Results

A MANOVA was conducted on the two dependent variables of the summated 10-point BOS and the overall rating. Significant multivariate effects (all $p<.01$) resulted for the main effects of base performance and serial position of the inconsistent performance, for the two-way interaction of base performance by serial position, and for the three-way interaction of delay by base performance by serial position.

Univariate effects on the summated BOS were significant for base performance [$F(1,544)=77.14$, $p<.01$] and the base performance by serial position interaction [$F(4,544)=3.50$, $p<.01$]. The main effect indicates that having good base performance is rated better than having poor base performance. The results of the interaction are of more interest and are graphed in Figure 1. Ratings of good base performance are lowest when an inconsistent performance occurs in second position. A recent poor performance (in fourth position) did not have detrimental effects when three good performances had preceded it. None of the means from the good base performance were significantly different, however, by Student-Newman-Keuls (SNK) multiple comparisons. Ratings of poor base performance generally mirrored those of the good base performance, although having a good performance in the series tended to raise the overall rating above the

control group of all four poor performances. An occurrence of good performance early on did not help as much as later occurences. SNK comparisons showed that the mean of the control group was significantly lower than those for the group who viewed a good performance second and the group who viewed a good performance fourth.

Univariate effects on the one-item overall rating were significant for base performance [$F(1,544)=716.16$, $p<.01$], serial position [$F(4,544)=7.05$, $p<.01$], the interaction of base performance by serial position [$F(4,544)=14.91$, $p<.01$], and the interaction of delay by base performance by serial position [$F(4,544)=4.86$, $p<.01$]. For interpretive purposes, this three-way interaction is graphed in Figure 2. For the immediate rating conditions, there is a tendency toward a recency effect in these overall impressions. The later the inconsistent performance occurs in the series, the greater the effect it seems to have. Student-Newman-Keuls (SNK) tests for differences in means indicated that for good base performance the ratings were lower than the control group if the poor performance occurred second or fourth. For poor base performance, the ratings were significantly higher than the control group when the good performance occurred last. For the delayed rating conditions, the results for the good base performance are similar to those for the summated BOS reported previously. There is a greater detriment for a single poor performance when it occurs in second or third position rather than anywhere else. SNK's showed the ratings for these two conditions to be significantly lower than the ratings in any of the three other delayed rating conditions with good base performance. For the poor base performance, any occurence of good performance does not seem help the

overall impression ratings.

Similar analyses were also conducted for the actions to take toward the instructor scale. A MANOVA on the three variables, counsel, reward, and punish, produced significant main effects for delay [$F(3, 539)=2.65$, $p<.05$], base performance [$F(3, 539)=109.13$, $p<.01$], and serial position [$F(12, 1426.35)=4.30$, $p<.01$]. Significant interactions were attained for base performance by serial position [$F(12, 1426.35)=4.57$, $p<.01$] and delay by base performance by serial position [$F(12, 1426.35)=3.22$, $p<.01$]. Univariate ANOVAs were done for each action separately. For the appropriateness to counsel, the overall F and the three main effect F's were all significant at the .01 level (see Table 1). The base performance by serial position interaction and the three-way interaction of delay by base performance by serial position were significant at the .01 and .05 levels, respectively (also in Table 1). For interpretation, the three factor interaction is graphed in Figure 3. SNK comparisons indicated that individuals in the good base performance control groups (both the delayed and the immediate) rated the appropriateness to counsel significantly lower than individuals assigned to any other condition. Therefore, having even one poor performance resulted in higher ratings for counseling. Individuals in the delayed condition with good base performance rated counseling significantly less appropriate if the single poor performance occurred last rather than in second position.

For the ratings of the appropriateness to punish, the overall F was significant [$F(19, 541)=18.05$, $p<.01$]. Significant effects resulted for the base performance main effect [F $(1, 541)=234.77$, $p<.01$], the interaction of

base performance and serial position [F (4, 541)=7.87, p<.Ø1], and the three factor interaction, delay by base performance by serial position [F (4, 541)=7.31, p<.Ø1]. The main effect for base performance indicates that punishment is viewed more appropriate for poor base performance than for good base performance. The three-way interaction is graphed in Figure 4. SNK comparisons indicated significantly lower punishment ratings by subjects in the immediate delay--poor base performance condition when a single good performance occurred last in the series relative to the control group. In the immediate delay--good base performance condition, subjects gave significantly higher punishment ratings when a single poor performance occurred last as compared to the control group. For the delayed condition with good base performance, subjects who viewed the inconsistent poor performance in the second and third serial positions rated punishment as more appropriate than subjects in the control group and the group who viewed it last. No other significant differences occurred within conditions.

Finally, for the ratings of the appropriateness to reward, the overall F (19, 541)=16.96, p<.Ø1, was significant. Main effects resulted for base performance [F(1, 541)=2Ø3.48, p<.Ø1] and serial position [F(4, 541)=7.38, p<.Ø1]. The interaction of base performance and serial position was the only other significant effect [F(4, 541)=7.25, p<.Ø1], and it is presented in Figure 5. SNK multiple comparisons showed that subjects in the good base performance conditions gave higher reward ratings than subjects in the poor base performance condtions. Subjects who only viewed good performances (control group) gave higher reward ratings than subjects who had seen even one poor performance. Subjects in the good base performance conditions who

saw poor performance either initially or last rated rewarding as more appropriate than subjects who viewed it second.

Discussion

The results for the BOS ratings would seem to indicate evidence for ideas presented by Schuh (1978) and Webster (1982). They hypothesized that contrast effects would occur when rating ambiguous performance, while assimilation would occur when rating extreme performance. The design of the current study focused on the extreme performance situation. No recency effect resulted for the inconsistent poor performance in a series of good performances. It would appear that a schema develops which biases memory for ratee behavior, as was found in Murphy et al. (1986). The detriment of having an instance of poor performance only resulted when it occurred in second position, before the schema was established, but this effect was not significant. When poor performance was the base, a good performance did improve ratings on the BOS. Raters generally prefer to give positive ratings (Landy & Farr, 1983), and the BOS asks for occurrences of specific behaviors, so the raters in the poor base performance did seem to take this good performance into account. This study did not directly test for the contrast versus assimilation effect; assimilation was inferred from the lack of a recency effect. Further research needs to test this directly by examining ratings of each incident of performance, not just overall ratings.

When rating using an overall impression rather than specific behaviors, the results looked somewhat different. Here there was perhaps no need to justify an overall unfavorable impression as is required in behavioral ratings such as BOS; hence, when viewing the lectures in the poor base

performance conditions across the period of a week, raters maintained their negative evaluation, unaffected by an incidence of good performance. If, however, they viewed the tapes consecutively and rated them immediately, the recent good performance caught their attention producing a recency effect, similar to the one reported in Steiner and Rain (in press).

The good base performance conditions produced results similar to the BOS ratings for the delayed conditions but not for the immediate. It is possible that a schema for level of performance develops over time in the delayed conditions. Thus, the initial poor performance is forgotten while the recent poor performance is subject to assimilation or attributed to unstable causes since the schema is well established at that point. It may be more difficult to establish a schema when the inconsistent performance occurs in the middle of the series; performance would perhaps appear more variable. In the immediate conditions, the attention decrement hypothesis would seem to be operating. The different recent performance captures the attention of the subject in this time period and results in greater weighting of the recent information.

In performance rating, raters seem to be unaffected by the occurrence of a single poor performance when it occurs early or late in a series of good performances. This finding is in opposition to the Steiner and Rain (in press) results where recency effects predominated. Considering both studies, we would conclude that raters who rate individual performances and are therefore attentive to each performance will tend give extra weight to the inconsistent performance when it occurs late in the series. On the other hand, raters who do not pay special attention to each performance are

likely to underattend to the recent poor performance due to waning attention. As in Steiner and Rain (in press), a recent good performance always seems to help. Further research is needed to understand why the ratings are affected when the inconsistent performance occurs at the second position. As mentioned earlier, it may be due to the difficulty of establishing a schema. Attribution theory (Kelley, 1973) would also seem to be a fruitful avenue to pursue to investigate whether raters attribute the inconsistent performance to internal or external factors and whether these attributions affect subsequent ratings. Attributing the inconsistent performance to external factors would presumably result in ratings that ignore the inconsistent performance; whereas attributions to internal factors would probably take the inconsistent performance into account.

With regard to actions to take toward the instructor, subjects recommended counseling and punishment as more appropriate for poor base performance and reward as more appropriate for good base performance. The interactions of various factors with serial position of inconsistent performance are of greater interest. For counseling, the major finding was that in the delayed condition with good performance, counseling is rated as less appropriate when poor performance occurs last. Raters seem willing to overlook this poor performance when performance has otherwise been good; perhaps they attribute it to unstable causes and therefore disregard it.

Results for punishment were more varied, perhaps because punishment is a stronger action than counseling. If base performance was poor and the delay was immediate, a recent good performance made punishment a less appropriate response than if all performances were poor. These results

paralleled those for the same condition on the overall ratings. Punishment was probably rated as less appropriate because performance was viewed as better. The ratings for punishment in the good base performance conditions was also similar to the results for overall ratings. In the immediate rating conditions, if poor performance was viewed last, punishment was viewed as more appropriate, just as performance was rated more favorably. And in the delayed conditions, when poor performance occurred in second or third position it was punished more, just as it was rated less favorably.

Finally, having a single poor performance resulted in lower reward ratings. Viewing poor performance second resulted in lower reward ratings than viewing it initially or last, reflecting the tendency to make lower overall ratings when poor performance occured second.

Similar explanations can be applied to the action scales as were relevant to the performance rating scales. The results for punishment more closely parallel the overall rating results than either the results for counsel or for reward. Punishment is not only a strong action to take, but it is also the action most closely tied to poor performance. Reviewing the absolute ratings of performance, subjects tended to rate all performace, even the good base performance, at about the average or lower level on the scale. Punishment would presumably be the action that would be both somewhat appropriate and most sensitive to different performance levels. Attribution theory may also prove useful in explaining the suggested actions.

References

Anderson, N. H. (1971). Integration theory and attitude change. Psychological Review, 78, 171-206.

Cooper, W. (1981). Ubiquitous halo. Psychological Bulletin, 90, 218-244.

DeNisi, A. S., Cafferty, J. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33, 360-396.

DeNisi, A. S., & Stevens, G. E. (1981). Profiles of performance, performance evaluations, and personnel decisions. Academy of Management Journal, 24, 592-602.

Feldman, J. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Kane, J. S., & Lawler, E. E. (1979). Performance appraisal effectiveness: Its assessment and determinants. Research in Organizational Behavior, 1, 425-478.

Kelley, H. (1973). The process of causal attribution. American Psychologist, 28, 107-128.

Landy, F., & Farr, J. (1980). Performance rating. Psychological Bulletin, 87, 77-107.

Landy, F. J., & Farr, J. L. (1983). The measurement of work performance. New York: Academic Press.

Luchins, A. S. (1957). Experimental attempts to minimize the impact of first impressions. In C. E. Houland (Ed.), The order of presentation in persuasion. New Haven, CT: Yale University Press.

Murphy, K. R., Balzer, W. K., Lockhart, M. C., & Eisenman, E. J. (1985). Effects of previous performance on evaluations of present performance. Journal of Applied Psychology, 70, 72-84.

Murphy, K. R., Gannett, B. A., Herr, B. M., & Chen, J. A. (1986). Effects of subsequent performance on evaluations of previous performance. Journal of Applied Psychology, 71, 427-431.

Murphy, K. R., Martin, C., & Garcia, M. (1982). Do behavioral observation scales measure observation? Journal of Applied Psychology, 67, 562-567.

Rambo, W. W., Chomiak, A. M., & Price, J. M. (1983). Consistency of performance under stable conditions. Journal of Applied Psychology, 68, 78-87.

Ronan, W. W., & Prien, E. P. (1971). Toward a criterion theory: A review and analysis of research and opinion. New York: Appleton-Century-Crofts.

Rothe, H. F. (1978). Output rates among industrial employees. Journal of Applied Psychology, 63, 40-46.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). Person Perception. Reading, MA: Addison-Wesley.

Schuh, A. (1978). Contrast effects in the interview. Bulletin of the Psychonomic Society, 11, 195-196.

Scott, W. E., & Hamner, W. C. (1975). The influence of variations in performance profiles on the performance evaluation process: An examination of the validity of the criterion. Organizational Behavior and Human Performance, 14, 360-370.

Steiner, D. D. & Rain, J. S. (in press). Immediate and delayed primacy and recency effects in performance evaluation. Journal of Applied Psychology.

Webster, E. C. (1982). The employment interview: A social judgment process. Schomberg, Ontario: SIP Publications.

FIGURE 1
BASE PERFORMANCE X SERIAL POSITION INTERACTION

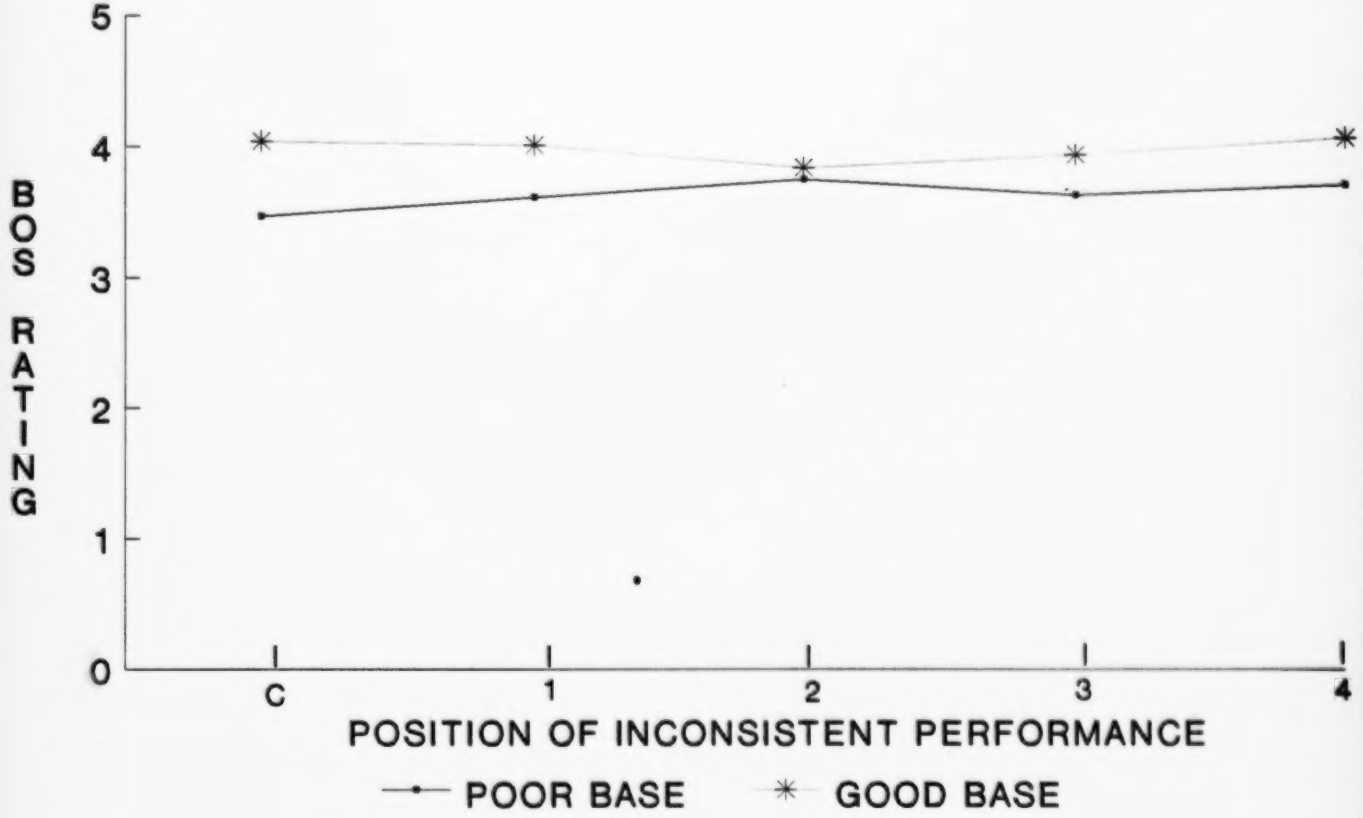

FIGURE 2
DELAY X BASE PERFORMANCE X SERIAL POSITION INTERACTION

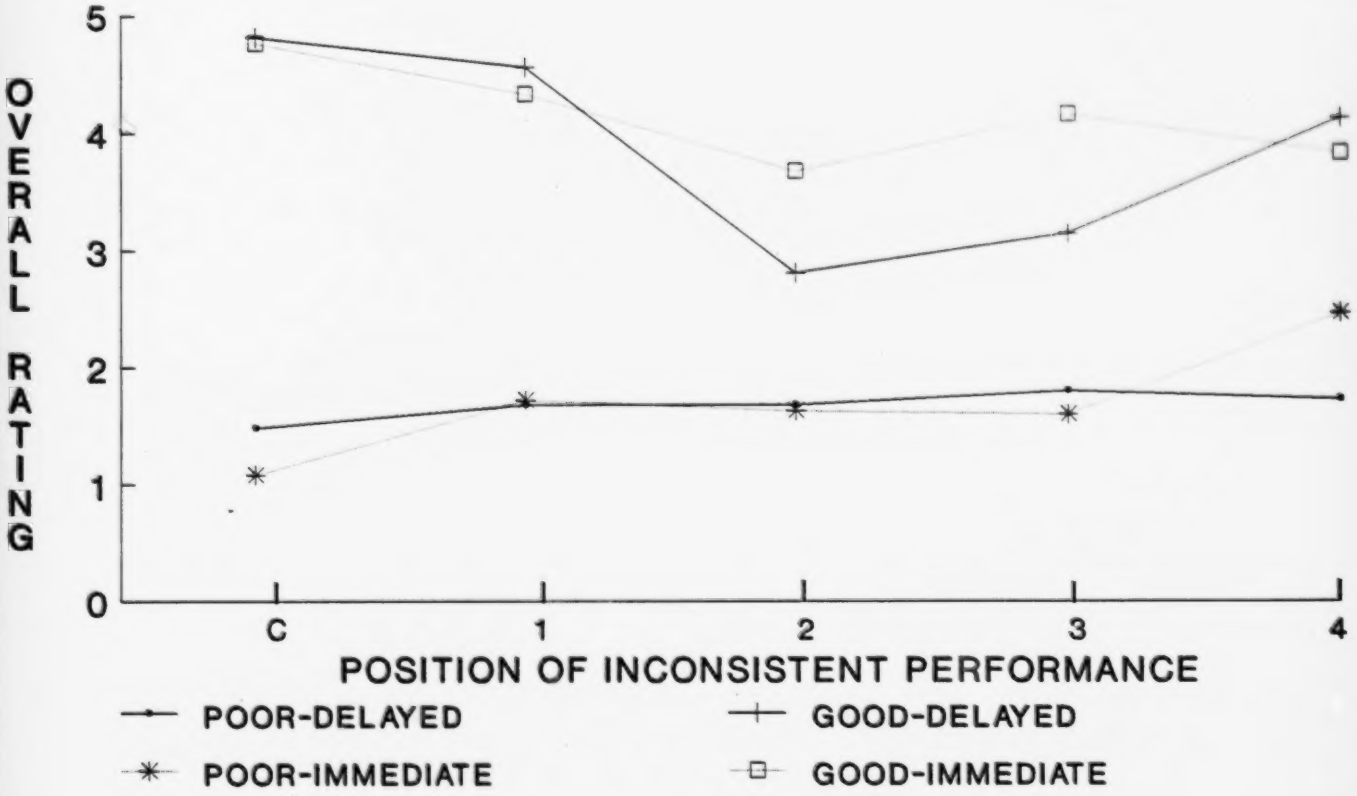

FIGURE 3
DELAY X BASE PERFORMANCE X SERIAL POSITION INTERACTION

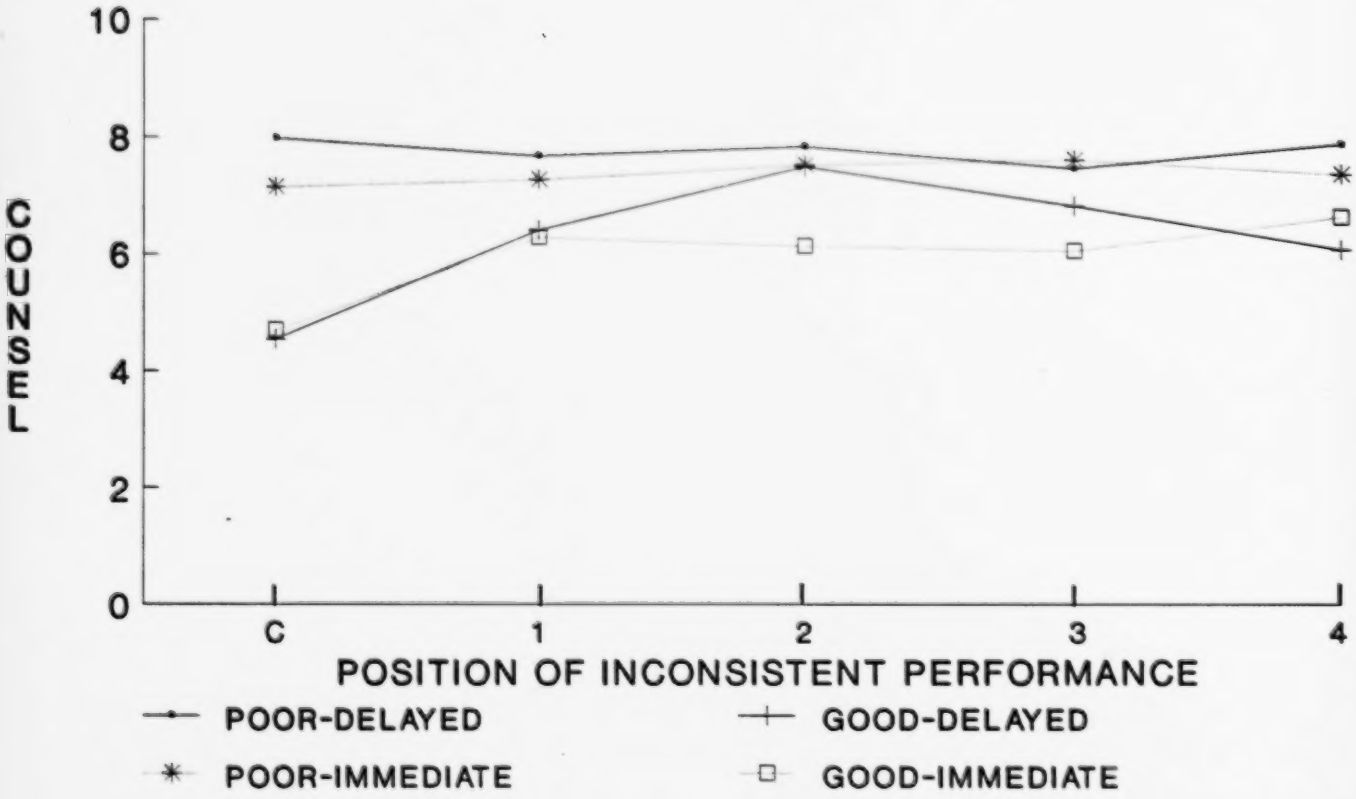

FIGURE 4
DELAY X BASE PERFORMANCE X SERIAL POSITION INTERACTION

FIGURE 5
BASE PERFORMANCE X SERIAL POSITION INTERACTION